MPAS-JEDI 3D/4DEnVar

I-Han Chen

ihanchen@ucar.edu

NCAR | MESOSCALE & MICROSCALE METEOROLOGY

Overview

1. **Variational Cost Function**
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

The Problem

We want to find the **analysis state** (x) that minimizing a cost function with an optimal fit to the background and observations.

$$J(x) = \boxed{\frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b)} + \boxed{\frac{1}{2}(h(x) - y)^T \mathbf{R}^{-1}(h(x) - y)}$$

Distance to background

Distance to observations

Incremental Cost Function in JEDI

Liu et al. (2022)

Full-form

$$J(x) = \frac{1}{2}(x - x_b)^T \boldsymbol{B}^{-1}(x - x_b) + \frac{1}{2}(h(x) - y)^T \boldsymbol{R}^{-1}(h(x) - y)$$

Incremental-form

$$J(\delta x) = \frac{1}{2}(\delta x - \delta x_g)^T \boldsymbol{B}^{-1}(\delta x - \delta x_g) + \frac{1}{2}(\boldsymbol{H}\delta x - d)^T \boldsymbol{R}^{-1}(\boldsymbol{H}\delta x - d)$$

$\boldsymbol{\delta x = x - x_g}$ $\delta x_g = x_b - x_g$ $\boldsymbol{d = y - h(x_g)}$

The minimization deals with increments to a known reference state

- Cost function minimizes δx=x− x_g instead of the full state (x)
- Start from $x_g = x_b$ and $\delta x_g = 0$
- **After minimization->** $x_a = x_g + \delta x$

Appropriately assign B and R is critical

We want to find the analysis state (x) that minimizing a cost function with **an optimal fit** to the background and observations.

Distance to background **Distance to observations**

$$J(\delta x) = \frac{1}{2}\left(\delta x - \delta x_g\right)^T B^{-1}\left(\delta x - \delta x_g\right) + \frac{1}{2}(H\delta x - d)^T R^{-1}(H\delta x - d)$$

The weighting between the two components is determined by B (background error) and R (observation error).

- A larger B means background is less accurate -> x will get closer to observation
- A larger R means observation is less accurate -> x will get closer to background

Two types of background error covariance (B)

$$J(\delta x) = \frac{1}{2}\left(\delta x - \delta x_g\right)^T B^{-1}\left(\delta x - \delta x_g\right) + \frac{1}{2}(H\delta x - d)^T R^{-1}(H\delta x - d)$$

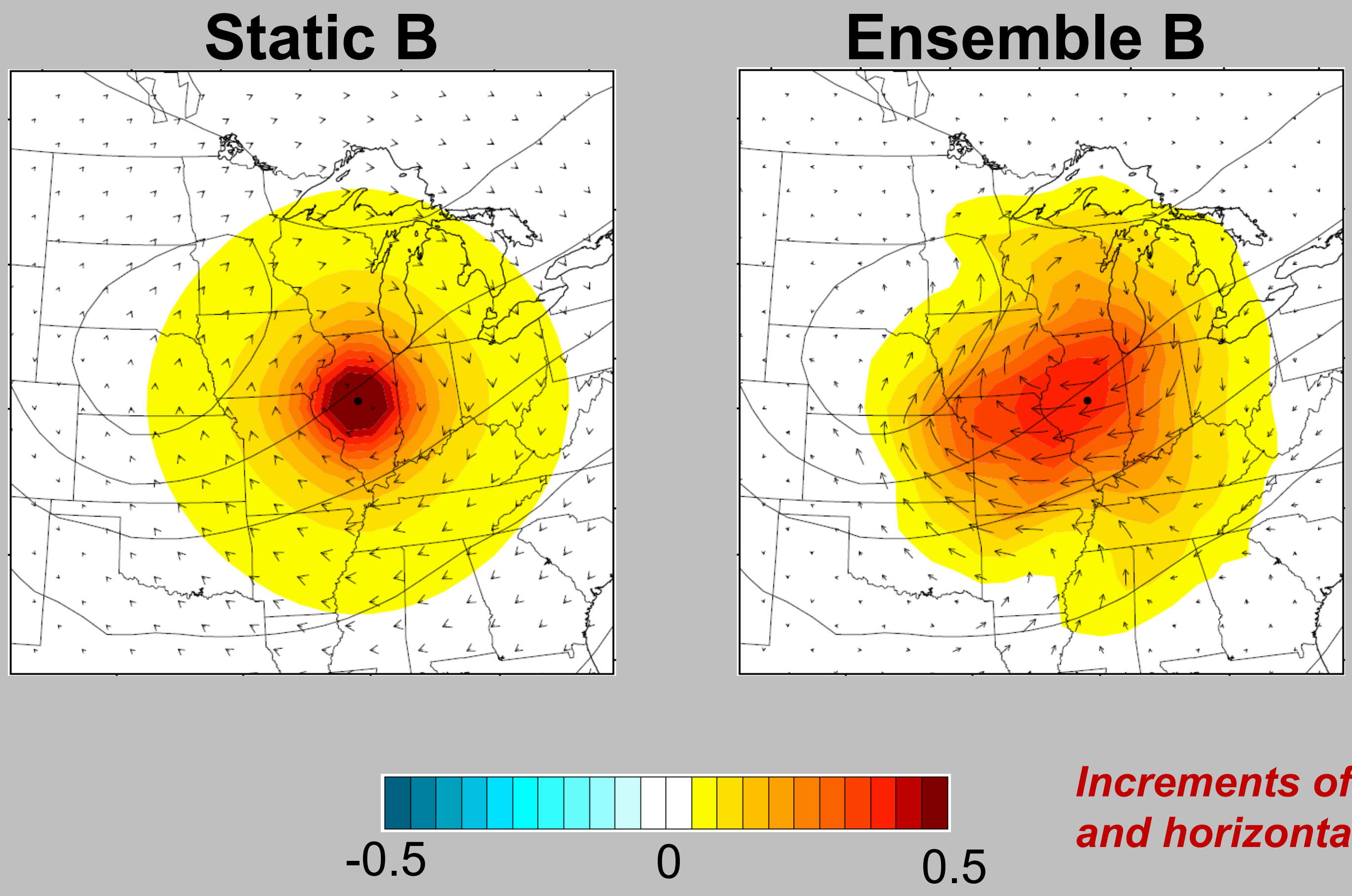
1. **Static B**

   -> from statistic, does not vary with time

2. **Ensemble B**

   -> flow-dependent, reflect the background error in different time

Example to show the B effect (Single observation tests)



Increments of temperature (shaded) and horizontal winds (vector)

Ensemble ***B***:

- Errors of the day are sampled
- flow-dependent update

Overview

1. Variational Cost Function
2. **Ensemble Error Covariance Matrix**
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

Derive B matrix from an ensemble of forecasts

ensemble size

State variable of ensemble mean

$$\boldsymbol{B_e} = \frac{1}{n-1}\sum_{i=1}^{\boldsymbol{n}}(\boldsymbol{x_i} - \boldsymbol{\overline{x}})(x_i - \bar{x})^T$$

State variable of each ensemble member

$$\boldsymbol{B_e} = \frac{1}{n-1}\sum_{i=1}^{\boldsymbol{n}}(\delta x_{\boldsymbol{i}})(\delta x_{\boldsymbol{i}})^T$$

ensemble perturbation

- The ensemble mean provides an estimation of the truth
- The perturbations from the mean estimate the uncertainty, which is used to model background-error covariance matrix.

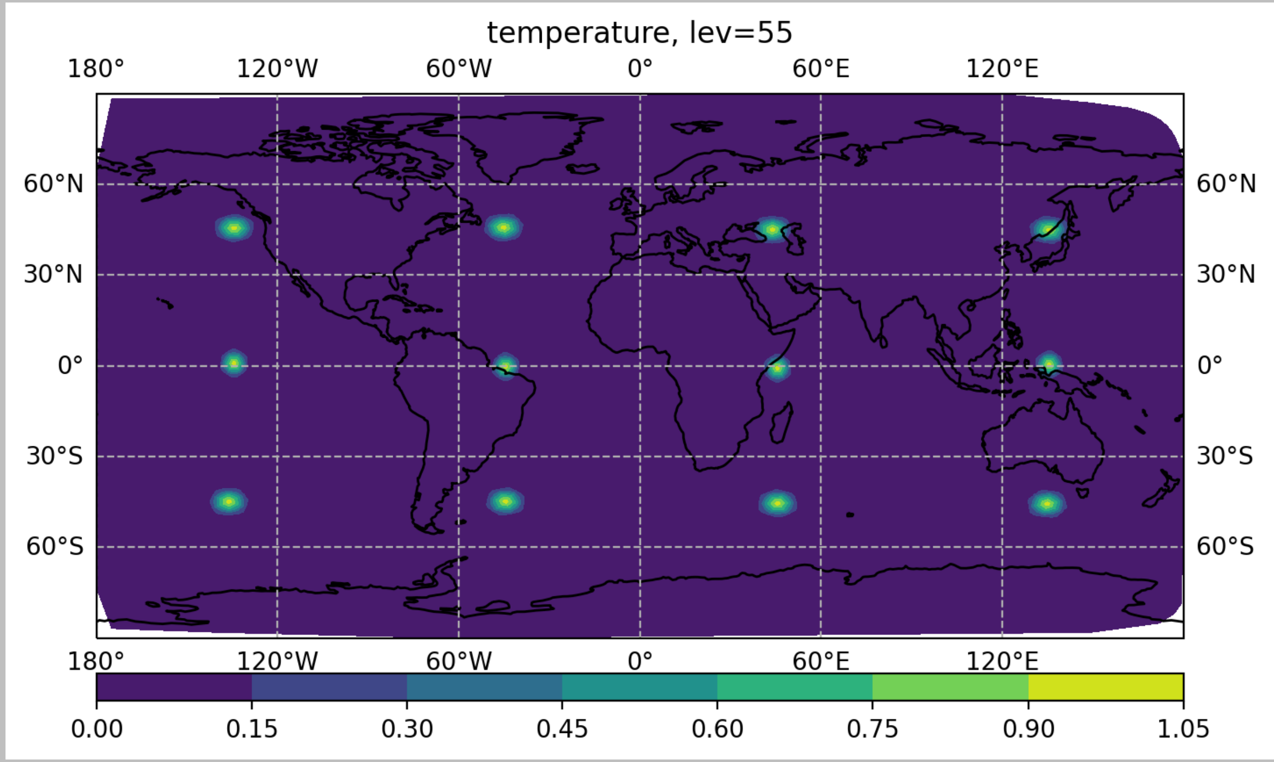
Localization of the *B* matix

Because we do not have a complete estimate of ***B*** (e.g., limited ensemble size) we need to use localization

Basic idea: observations should only influence an area nearby the observation

$$\boldsymbol{B} = \boldsymbol{L} \circ \boldsymbol{B_e}$$

Small localization



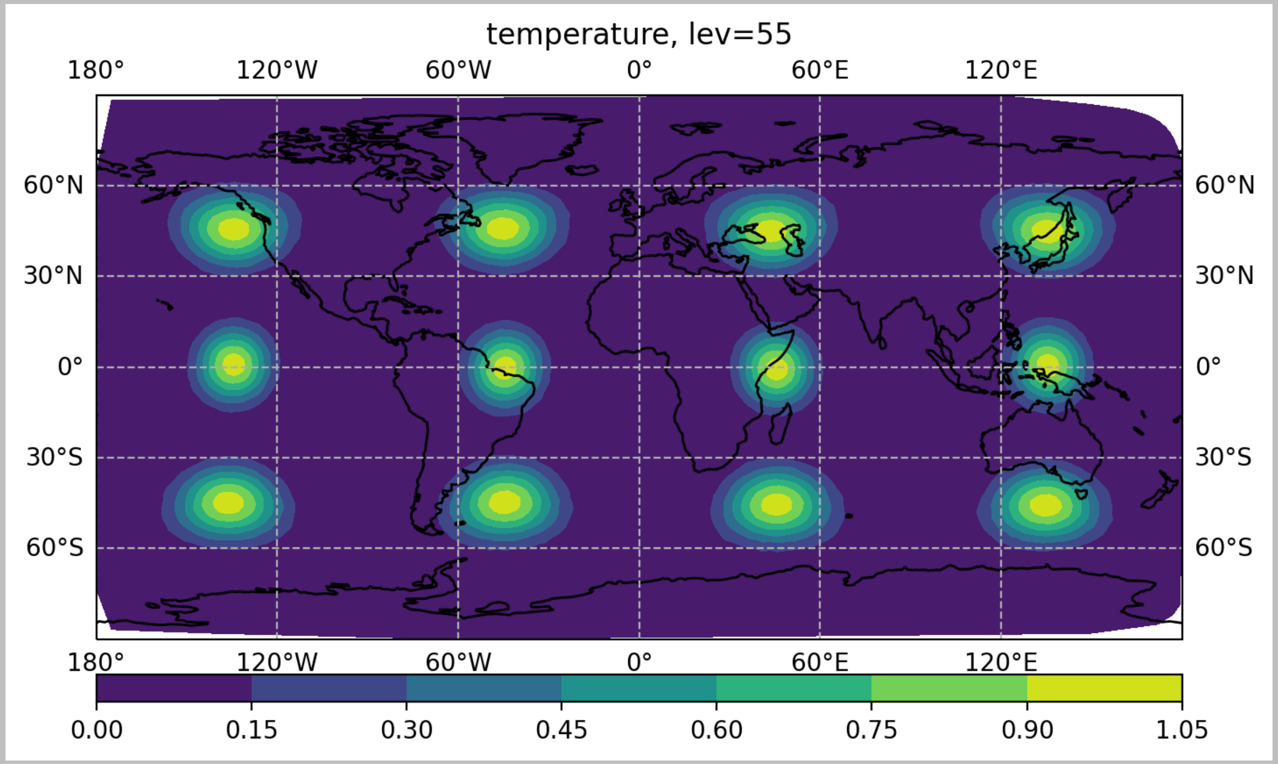
Localization of the *B* matix

Because we do not have a complete estimate of ***B*** (e.g., limited ensemble size) we need to use localization

Basic idea: observations should only influence an area nearby the observation

$$\boldsymbol{B} = \boldsymbol{L} \circ \boldsymbol{B}_e$$

large localization



Benefits of using an ensemble to estimate *B*

- Simple to implement

- Provides a flow-dependent estimate of the errors and uncertainties
  - Depends on the quality of the ensemble

- Incorporates ensemble estimate of background errors within the variational update
  - **Still updates a deterministic forecast**

EnVar uses a pure ensemble B to updates a deterministic forecast

In hybrid methods, B can be a weighting sum between static B ($\boldsymbol{B_s}$) and ensemble B ($\boldsymbol{B_e}$).

$$\boldsymbol{B} = \beta_s \boldsymbol{B_s} + \beta_e \boldsymbol{B_e}$$
$$\beta_s + \beta_e = 1$$

=1

pure ensemble B

Overview

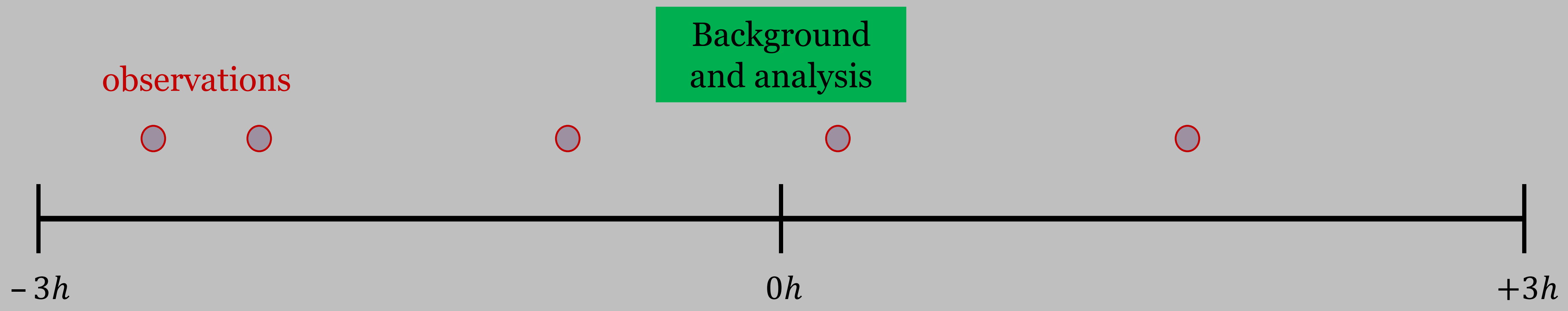
1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. **Overview of 3DEnVar**
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

3DEnVar

$$J(x) = \frac{1}{2}(x - x_b)^T \boldsymbol{B}^{-1}(x - x_b) + \frac{1}{2}(h(x) - y)^T \boldsymbol{R}^{-1}(h(x) - y)$$

- We assume that all observations $\boldsymbol{y_o}$ are valid at the same time.
- Usually valid at the center of the window (i.e. at the same time as $\boldsymbol{x}$ and $\boldsymbol{x_b}$)

3DEnVar using a 6h assimilation window



- All observations in 3DEnVar are assumed to be valid at the same time as the background

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. **Setting up a .yaml file for 3DEnVar**
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

Configure the analysis time for 3DEnvar

```
_member config: &memberConfig
  date: &analysisdate '2018-04-15T00:00:00Z'
  state variables: &incvars
  - temperature
  - spechum
  - uReconstructZonal
  - uReconstructMeridional
  - surface_pressure
  stream name: ensemble
cost function:
  cost type: 3D-Var
  window begin: '2018-04-14T21:00:00Z'
  window length: PT6H
  geometry:
    nml_file: "./Data/480km/namelist.atmosphere_2018041500"
    streams_file: "./Data/480km/streams.atmosphere"
    deallocate non-da fields: true
  analysis variables: *incvars
  background:
    state variables: [temperature, spechum, uReconstructZonal, uReconstructMeridional, surface_pressure,
                      theta, rho, u, qv, pressure, landmask, xice, snowc, skintemp, ivgtyp, isltyp,
                      snowh, vegfra, u10, v10, lai, smois, tslb, pressure_p]
    filename: "./Data/480km/bg/restart.2018-04-15_00.00.00.nc"
    date: *analysisdate
```

analysis time (center of window)

Start of assimilation window
length of assimilation window

First guess (should be at analysis time)

Configure the ensemble B

```yaml
background error:
  covariance model: ensemble
  localization:
    localization method: SABER
    saber central block:
      saber block name: BUMP_NICAS
      active variables: *incvars
      read:
        io:
          files prefix: Data/bump/mpas_parametersbump_loc
        drivers:
          multivariate strategy: duplicated
          read local nicas: true
  members:
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
```

set ensemble *B* for 3DEnVar

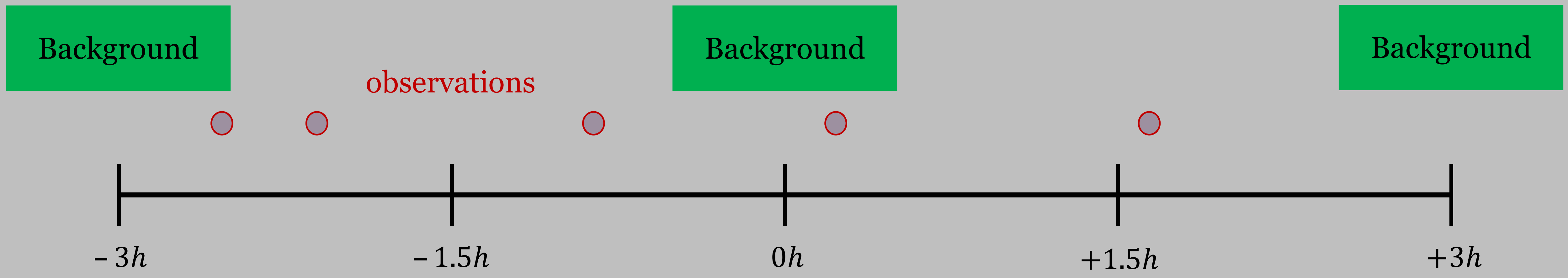
Specifying members used to compute ensemble B

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. **Overview of 4DEnVar**
6. Setting up a .yaml file for 4DEnVar

4DEnVar

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}\sum_{k=1}^{K}(Hx_k - y_k)^T {R_k}^{-1}(Hx_k - y_k)$$

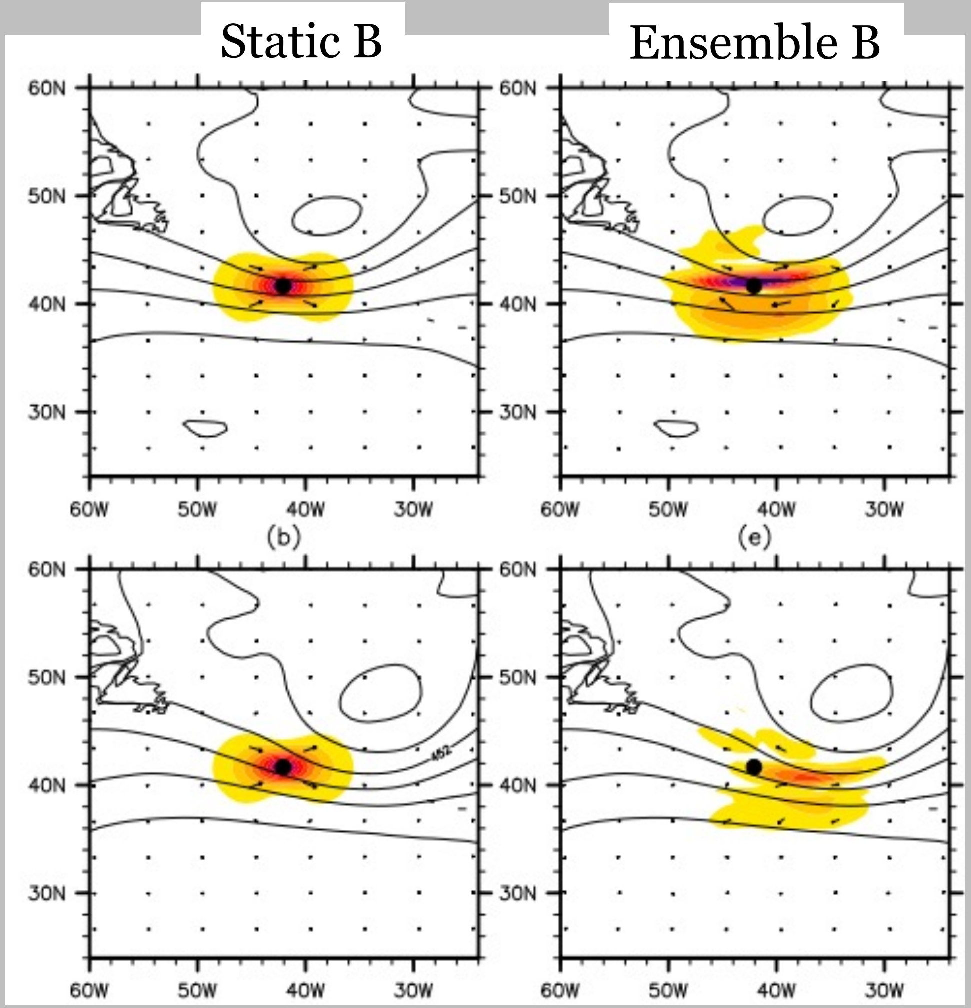


- All observations in 4DEnVar are binned within a smaller subwindow and innovations $(Hx - y_o)$ are calculated relative to background valid at that time.
- Ensemble needed at the center of each subwindow (***K*** ensemble required).

The 4D ensemble *B* is used to propagate the innovation

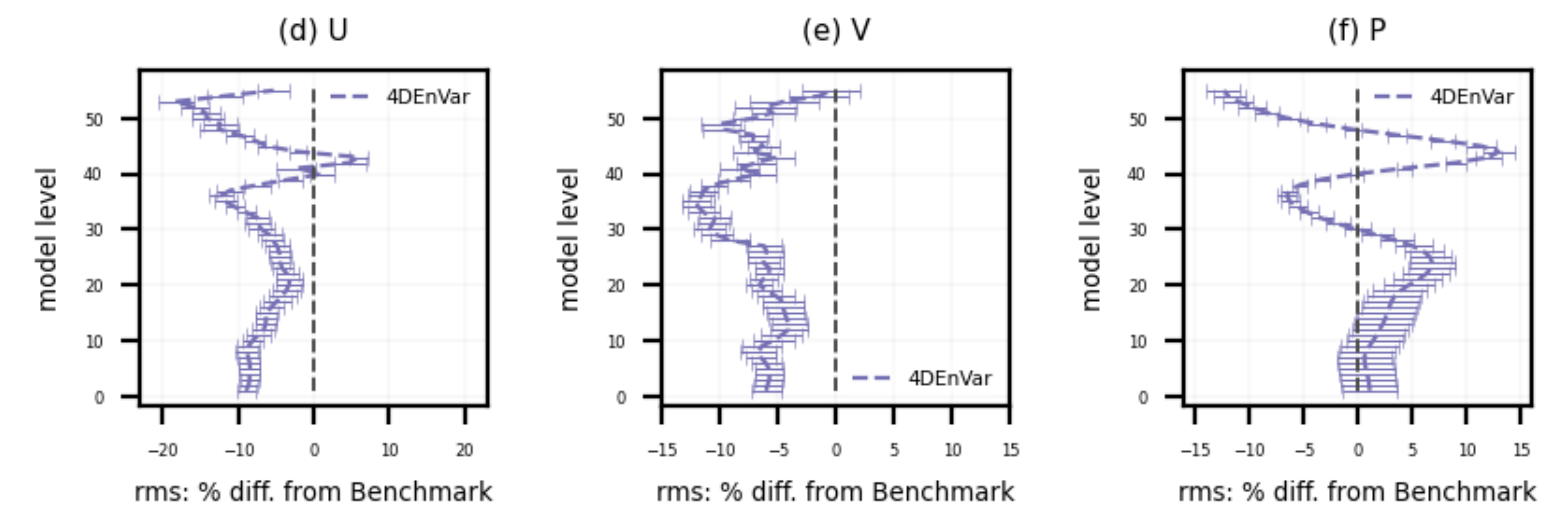
Start of window

end of window



Lorenc et al. (2015)

4DEnVar vs 3DEnVar Performance



- Assimilating observations at their appropriate time improves the analyses

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. **Setting up a .yaml file for 4DEnVar**

Configure the analysis times for 4DEnvar

```yaml
_member config 1: &memberConfig1
  date: &date1 '2018-04-14T21:00:00Z'
  state variables: &incvars
  - temperature
  - spechum
  - uReconstructZonal
  - uReconstructMeridional
  - surface_pressure
  stream name: ensemble
_member config 2: &memberConfig2
  <<: *memberConfig1
  date: &date2 '2018-04-15T00:00:00Z'
_member config 3: &memberConfig3
  <<: *memberConfig1
  date: &date3 '2018-04-15T03:00:00Z'
cost function:
  cost type: 4D-Ens-Var
  window begin: '2018-04-14T21:00:00Z'
  window length: PT6H
  subwindow: PT3H
```

subwindow1

subwindow2

subwindow3

Background needed for each subwindow

```
cost function:
  cost type: 4D-Ens-Var
  window begin: '2018-04-14T21:00:00Z'
  window length: PT6H
  subwindow: PT3H
  geometry:
    nml_file: "./Data/480km/namelist.atmosphere_2018041500"
    streams_file: "./Data/480km/streams.atmosphere"
  analysis variables: *incvars
  background:
    states:
    - state variables: &stvars
                       [temperature, spechum, uReconstructZonal, uReconstructMeridional, surface_pressure,
                       theta, rho, u, qv, pressure, landmask, xice, snowc, skintemp, ivgtyp, isltyp,
                       snowh, vegfra, u10, v10, lai, smois, tslb, pressure_p]
      filename: "./Data/480km/bg/restart.2018-04-14_21.00.00.nc"
      date: *date1
    - state variables: *stvars
      filename: "./Data/480km/bg/restart.2018-04-15_00.00.00.nc"
      date: *date2
    - state variables: *stvars
      filename: "./Data/480km/bg/restart.2018-04-15_03.00.00.nc"
      date: *date3
```

bg (subwindow 1)

bg (subwindow 2)

bg (subwindow 3)

Configure the ensemble B

```
background error:
  covariance model: ensemble
  localization:
    localization method: SABER
    saber central block:
      saber block name: BUMP_NICAS
      active variables: *incvars
      read:
        io:
          files prefix: Data/bump/mpas_parametersbump_loc
        drivers:
          multivariate strategy: duplicated
          read local nicas: true
```

set ensemble *B* for 4DEnVar

Member file needed for each subwindow

```yaml
members:
- states:
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
```

Further reading: an alternative way to specify the *B* ensemble

```yaml
- covariance:
    covariance model: ensemble
    localization:
      localization method: SABER
      saber central block:
        saber block name: BUMP_NICAS
        active variables: *incvars
        read:
          io:
            files prefix: Data/bump/mpas_parametersbump_loc
          drivers:
            multivariate strategy: duplicated
            read local nicas: true
    members from template:
      template:
        <<: *memberConfig
        filename: Data/480km/bg/ensemble/mem%iMember%/x1.2562.init.2018-04-15_00.00.00.nc
      pattern: '%iMember%'
      start: 1
      zero padding: 2
      nmembers: 4
      except: [1] # chooses 2, 3, 4, 5
```

You can find it at
./mpas_bundle_v2/build/mpas-jedi/test/testinput /eda_3dhybrid_1.yaml

References

- Liu, Z., and Coauthors, 2022: Data assimilation for the Model for Prediction Across Scales - Atmosphere with the Joint Effort for Data assimilation Integration (JEDI-MPAS 1.0.0): EnVar implementation and evaluation. *Geosci. Model Dev.*, **15**, 7859–7878, https://doi.org/10.5194/gmd-15-7859-2022.
- Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weather Rev.*, **143**, 212–229, https://doi.org/10.1175/MWR-D-14-00195.1.